**Step 1:** Tải tập dữ liệu về máy: diabetes.csv

Dataset lấy từ Kaggle bộ dữ liệu này gồm có:

* **diabetes.csv**: Dữ liệu gốc.

**Link dataset:** <https://www.kaggle.com/code/fatihmehmetkoc/svc-79-suscesfully/input>

**Step 2:** Truy cập **svc-79-suscesfully** để lấy code đã được xây dựng sẵn trên Kaggle

Link:

<https://www.kaggle.com/code/fatihmehmetkoc/svc-79-suscesfully/notebook>

Nhấn vào “Download” để tải notebook về máy. Mở file notebook này bằng **Visual Studio Code** hoặc **Jupyter Notebook** để chỉnh sửa và thực hiện bước tiếp theo.

Step 3: Lấy code và thực hiện chạy chương trình trên Jupyter Notebook

Step 4: Copy đường dẫn tới file tập dữ liệu **diabetes.csv** (trong máy Local – máy tính mà bạn đang sử dụng trực tiếp, ở mục đã tải tập dữ liệu về) vào function (hàm) read\_csv. Ghi chú: read\_csv là một hàm trong thư viện **pandas** của Python

Ví du:

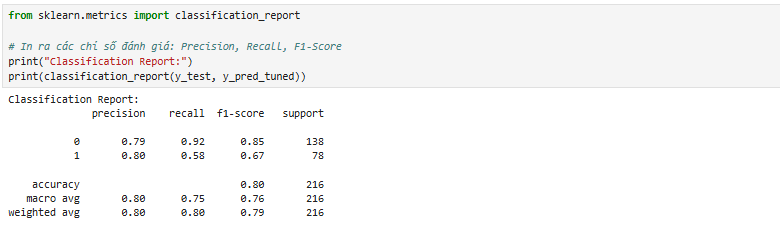
*# Nhập thư viện pandas, gọi tắt là pd*

import pandas as pd

*# Đọc dữ liệu từ file CSV*

data=pd.read\_csv("C:/Year4/KhaiPhaDuLieu/diabetes.csv")

**Step 5:**  kết quả các chỉ số đánh giá mô hình như độ chính xác (accuracy), độ nhạy (recall), độ đặc hiệu (precision), và điểm F1.



Hoàn tất quy trình tải dữ liệu, chuẩn bị môi trường và chạy chương trình đánh giá mô hình dự đoán nguy cơ bệnh tiểu đường với độ chính xác cao.